(1)

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

- 1a. REPORT SECURITY CLASSIFICATION: UNCLASSIFIED
- 1b. RESTRICTIVE MARKINGS:
- 2a. SECURITY CLASSIFICATION AUTHORITY:
- 2b. DECLASSIFICATION/DOWNGRADING SCHEDULE:
- 3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution is unlimited.
- 4. PERFORMING ORGANIZATION REPORT NUMBER(S):
- 5. MONITORING ORGANIZATION REPORT NUMBER(S):
- 6a. NAME OF PERFORMING ORGANIZATION: Naval Ocean Systems Center
- 6b. OFFICE SYMBOL (If applicable): NOSC
- 6c. ADDRESS (City, State and ZIP Code): San Diego, California 92152-5000
- 7a. NAME OF MONITORING ORGANIZATION:
- 7b. ADDRESS (City, State and ZIP Code):
- 8a. NAME OF FUNDING/SPONSORING ORGANIZATION: Office of Naval Technology
- 8b. OFFICE SYMBOL (If applicable): ONT
- 8c. ADDRESS (City, State and ZIP Code): Office of Chief of Naval Research, Arlington, VA 22217
- 9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER:
- 10. SOURCE OF FUNDING NUMBERS:

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	AGENCY ACCESSION NO.
0603131M	CC56	C3125	DN888 554

- 11. TITLE (Include Security Classification): VERY LOW DATA RATE VOICE COMMUNICATION
- 12. PERSONAL AUTHOR(S): R. Johnson
- 13a. TYPE OF REPORT: presentation/paper
- 13b. TIME COVERED: FROM TO
- 14. DATE OF REPORT (Year, Month, Day): September 1989
- 15. PAGE COUNT:
- 16. SUPPLEMENTARY NOTATION:
- 17. COSATI CODES:

FIELD	GROUP	SUB-GROUP

- 18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number): communications, computers, marcorps
- 19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The domain of very low data rate voice communications is not universally defined. For this discussion, the domain will be verbal communication at a maximum serial rate of a few hundred bits per second. Systems are ordered or under development which operate at 400 bps or less.

Some areas of communication have a desperate need for very low rates. This compelling need arises from the laws of physics and the mathematical relationships inherent in the applications. Because of its traditionally higher data rate, digital voice hasn't been a practical choice for some of these areas until recently.

I'll briefly describe a group of these applications in general and then cover the details in the next section. One application is using digital voice on narrow-band radio channels (as in the typical 3kHz wide single sideband voice channel in the VHF bands). Another is jam-resistance where massive redundancy or unique operating techniques require a low rate. A third use is balancing the relative load in integrated voice/data systems. A related application is in disguising a voice channel as a data channel. A final application is in multiplexing of several users onto one channel.

Published in *Proceedings of Speech Tech '87*. April 1987.

- 20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: [X] UNCLASSIFIED/UNLIMITED [ ] SAME AS RPT [ ] DTIC USERS
- 21. ABSTRACT SECURITY CLASSIFICATION: UNCLASSIFIED
- 22a. NAME OF RESPONSIBLE INDIVIDUAL: R. Johnson
- 22b. TELEPHONE (Include Area Code): (619) 553-3653
- 22c. OFFICE SYMBOL: Code 441

DD FORM 1473, 84 JAN — 83 APR EDITION MAY BE USED UNTIL EXHAUSTED. ALL OTHER EDITIONS ARE OBSOLETE

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

THE OFFICIAL PROCEEDINGS OF

SPEECH TECH'87

VOICE INPUT/OUTPUT APPLICATIONS SHOW AND CONFERENCE

Ralph

April 28-30, 1987
Waldorf-Astoria Hotel
New York City

Published by
MEDIA DIMENSIONS, INC.
Producers of SPEECH TECH '87
and

SPEECH TECHNOLOGY
MAN MACHINE VOICE COMMUNICATIONS

VERY LOW DATA RATE VOICE COMMUNICATION

Ralph Johnson

Naval Ocean Systems Center
San Diego, CA 92152-5000

INTRODUCTION

The domain of very low data rate voice communications is not universally defined. For this discussion, the domain will be verbal communication at a maximum serial rate of a few hundred bits per second. Systems are ordered or under development which operate at 400 bps or less.

Some areas of communication have a desperate need for very low rates. This compelling need arises from the laws of physics and the mathematical relationships inherent in the applications. Because of its traditionally higher data rate, digital voice hasn't been a practical choice for some of these areas until recently.

I'll briefly describe a group of these applications in general and then cover the details in the next section. One application is using digital voice on narrow-band radio channels (as in the typical 3kHz wide single sideband voice channel in the VHF bands). Another is jam-resistance where massive redundancy or unique operating techniques require a low rate. A third use is balancing the relative load in integrated voice/data systems. A related application is in disguising a voice channel as a data channel. A final application is in multiplexing of several users onto one channel.

CONSTRAINTS WHICH DEMAND LOW RATE

The compelling need for low data rate has its basis in the laws of nature. It shouldn't be confused with the apparent (but temporary) limits in speeds we have historically seen in other electronics systems. This need for low rates isn't caused by limitations of hardware speed or price and won't be eliminated by new technology.

These constraints which are due to nature or to the use of unique techniques include:

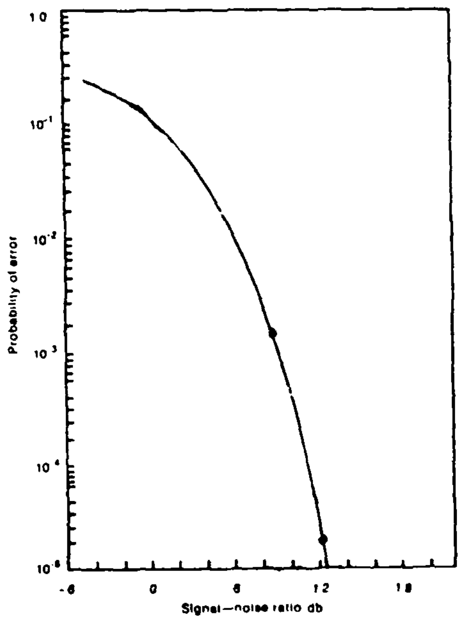
A. Restricted bandwidth - Single sideband or telephone voice channels have bandwidths on the order of 3kHz. Shannon's Theorom gives the maximum theoretical channel capacity. Stated simply, information capacity is proportional to bandwidth and to the signal-to-noise ratio (SNR).

Shannon's Law : $C = BW \times \log2 (1 + SNR)$

Radio systems are always expected to operate at minimal SNR's so bandwidth is the usual "variable" which limits the capacity of a radio channel. For a 10 dB SNR and a 3kHz bandwidth, Shannon's maximum capacity computes to 30k bps. Unfortunately, considerably less is achieved in reality. A rule of thumb is one bit per hertz over a consistent channel. I've seen a commercially produced militarized radio modem which operates at 2400 bps.

Radio channels also suffer from fading and noise. These effects can be reduced by redundancy for error correction. The added "overhead" data bits cause a further reduction in the maximum allowable base data rate.

B. Noise on the channel - In order to receive a signal dependably, it must be consistently greater than the noise. The effects of noise on a received digital signal is usually shown as a Bit-Error-Rate curve. Usually the probability of a bit being in error is plotted on the vertical axis and signal to noise ratio is plotted on the horizontal axis. This is a special form of SNR which is the energy per bit compared to receiver noise. An example of such a curve is shown in the graph below.



BER PERFORMANCE

The typical curve shows that for high SNR's there is only a miniscule chance of error. However, at lower levels, drops in SNR can quickly produce unacceptable errors. For example at 12 dB the error rate is 1/100,000 while halving the power (3dB drop) shows the 9 dB error rate less than 1/1000. We see more than 100 times the errors when we cut the power in half. Looking at it another way, doubling the power per bit produces a 100 times improvement.

There are two ways to double the power per bit. One is to double the transmitter power, the other is to double the length of the bit time (and therefore its total power). Naturally transmitter power is usually already at a maximum for any given application. The best choice is to double the bit length, which requires cutting the bit rate in half. From

this we can easily see the pressure for lower bit rates brought about by physical laws.

C. Use of anti-jamming techniques - Jamming is a special case of noise. It is maliciously applied noise which requires a very strong signal to overcome. This alone favors the use of the low rate, high energy-per-bit systems, but special features of some anti-jamming techniques also require low rates.

One of the best approaches to jamming is to escape it. One escapist type of spread spectrum technique is called frequency hopping and it involves pseudo-random frequency shifts by the transmitter and receiver simultaneously. They both dwell on each "random" channel for a specified time. "Following" jammers attempt to track the shifts and move with them. They may be able to obliterate part of the dwell time at each new frequency. A low rate would minimize the data lost during such an overlap. A "fast hopper" shifts several times for each bit. A low data rate keeps the required hopping rate manageable. Slow changing data with respect to synchronization "jitter" is also desirable in all these hopping schemes.

D. Short channel access time - Some applications operate in a burst type communications environment. This means sending short blocks at high data rates but spaced at relatively long intervals. Naturally a very low base data rate allows a reasonable transfer of information during these brief "pop-ups".

A somewhat different version of this is the GTE work on meteor burst communication. In this case channels become available for a few seconds at sporadic times, and several seconds of voice must be sent in one short burst. By keeping the voice data rate well below 100 bps, they are able to communicate on this uncertain channel.

E. Multiplexing channel users - Several users can time share a given channel capacity. Lower data rates allow more users. Typically the benefits for radio channels are only half as great as they appear because two one-directional channels are needed for duplex communication. This just requires an even lower data rate.

F. Integrated voice and data - Voice data can overwhelm text data and overload an integrated system. Here is an example of why lower voice data rates are needed. : " This short sentence has six words" : It takes about three seconds to say this sample sentence. As ASCII text it has about 230 bits, or around 70 bits/sec. Speech data is very redundant with respect to semantic information. As typical digitized speech it would take thousands of bits/sec to deliver the same information content. A low rate can allow voice traffic on an integrated system with a load more in line with its information content.

G. Disguising voice traffic - Most current digital voice systems operate at thousands of bits per second. For this reason, any channel operating at only a few hundred bits per second can be assumed to be data. Unfortunately, monitoring normal patterns of voice-to-data ratios and comparing them to any current activity can provide intelligence information. Low rate voice can eliminate the distinction.

APPROACHES TO VERY LOW DATA RATE VOICE

Users want speaker independent, unlimited vocabulary communications with excellent understandability. At the same time they want very low data rates. Exactly what is wanted is not yet available, but progress is being made. We'll briefly consider the magnitude of the problem before we explore the approaches taken to solve it.

Telephone quality speech (4kHz voice bandwidth) requires an 8 kHz sampling rate (the Nyquist rate). Quantizing at 8 bits per sample yields a "raw" data rate of 64k bits per second. This is, in fact, the current standard rate for telephone communications, although a new 32k bit standard has recently been approved. Condensing this information to 400 bps requires a compression ratio of 160 to one.

The need for this huge reduction led to the following approaches:

A. LPC - Linear Predictive Coding - an attempt to mathematically model the speech creation process and send only the barest essentials needed to recreate it. There is a commercial product which is very understandable at 2400 bps, and there is a NATO standard for 2400 bps systems.

Some schemes start with LPC and attempt to quantize it or further process it in new ways. Quantizing consists of making a table of distinct values spanning all those likely to occur in speech. Each of these table entries consists of numerous bits describing the "quantized" value completely. Each LPC description of user speech is compared to the table and is given the nearest value. Low data rate results from sending only the few bits for a pointer into an identical copy of the table at the receiver. The numerous bits retrieved from the table entry are then used to reconstruct the speech.

Another mathematical reduction at Rockwell exploits the basic properties of spectrum representation by LPC. "Line spectrum pairs" are found giving minimal formant information. These pairs are quantized to the nearest neighbors to produce very low data rate. Reconstruction of speech requires this formant spectrum information along with pitch and gain. The latter are relatively slow-changing and can be represented with reasonably low bit count, allowing an overall low rate.

The disadvantage of these LPC approaches to date is that they have produced poor sound quality. They also can become speaker dependent (custom tables for each user). Their great advantage is that they allow an unlimited vocabulary, unlike another solution to be discussed later.

B. Segmentation - divides the speech up into optimal short blocks. These blocks are typically related to pitch periods. This too is a quantization approach with a "code book" of possible segments. The code book pointers are sent and the speech restructured from an identical code book. This is very similar to

the LPC approach except the segmentation algorithm is an important component of the plan. The same advantages and problems apply.

BBN and Lincoln Labs have used this approach. Lincoln Labs built hardware which operated below 1000 bps. I have heard a simulation of BBN's under-300 bps algorithm and found its performance impressive.

C. Recognized words - A speech recognizer is used to identify a user's words or phrases. A dictionary entry number for each word is sent and digital recording (or, less desirably a speech synthesizer) reconstructs the speech. A 1000 word vocabulary can yield about 20 bits per second. GTE is building a 1000 word-or-phrase system for meteor burst communication. A previous NOSC project produced a 200 phrase system using older technology.

The advantages of the recognized word scheme are the very lowest data rate and the high quality recorded sound possible. The drawbacks are inherent to recognizers. Such a system is speaker dependent, has a restricted vocabulary, faces the problem of recognition error recovery, and must contend with different "problem" words for each speaker as well as "problem" people (goats).

D. Recognized phonemes - the string of phonemes is extracted from the speech stream. These phonemes, along with pitch and duration information are sent to a resynthesizer. At a speech rate of 10 to 12 phonemes per second, a "raw" data rate of 160 to 200 bps can be expected. A lower theoretical limit for phoneme-only information (without triphone coding etc.) is about 70 bps. You may recall that this is about the rate we calculated for ASCII transmission of our sample sentence in the earlier example.

Advantages of this low rate system include: unlimited vocabulary, easy "recovery" from misrecognized phonemes (no syntax tree constraints and a human integrator), and compatibility with remote computer speech recognition. The drawbacks are speaker dependence (at present) and the lesser sound quality of a synthesizer.

NOSC VERY LOW RATE EFFORTS

A previously completed project used a now-obsolete recognizer and a 200 word vocabulary. It showed the communications advantages of very low rates but the vocabulary was too small and error recovery was unacceptable.

Currently we are beginning development on a phoneme based system. We have acquired a phonetic recognizer from Speech Systems Incorporated which we plan to use to identify phonemes along with their pitch, duration, and amplitude. We are scheduled to do a one-way demo this fiscal year. This is a two year project with two-way communication as the end product. We have promised 300 bps and expect to operate well under 200 bps.

FUTURE OF VERY LOW RATE VOICE

The demand for very low data rate voice is based on laws of physics. While new technology and algorithms can provide better performance at low rate, they can't eliminate the basic need. As long as channels are narrow, noisy or crowded there will be a need for very low data rate voice. We can expect the defined limit of very low data rate to drop as practical systems are developed, and we certainly can expect a great lowering in the ratio of voice-data to text-data for a given message.